Evaluating the Impact of Small Group Supplemental Math Enrichment in Kindergarten

Robin Jacob and Anna Erickson

University of Michigan

Shira Mattera

MDRC

Author Note

Robin Jacob, Institution for Social Research, University of Michigan; Anna Erickson, Institute for Social Research, University of Michigan; Shira Mattera, MDRC.

Making Pre-K Count and High 5s were developed by MDRC and the University of Michigan as part of the Robin Hood Early Childhood Research Initiative, which was established to identify and rigorously test promising early childhood interventions. The initiative is a partnership between Robin Hood, one of New York City's leading antipoverty organizations, and MDRC. Making Pre-K Count and High 5s are also supported with lead funding from the Heising-Simons Foundation, the Overdeck Family Foundation, and the Richard W. Goldman Family Foundation. The programs were conducted in partnership with Bank Street College, which implemented the coaching for teachers and facilitation of High 5s clubs; RTI International; and the New York City Department of Education and Administration for Children's Services.

Abstract

This article reports findings from a one-year evaluation of a kindergarten math enrichment program. A randomized design was used to assess the impact of High 5s on children's math skills, attitudes towards math, language ability, and executive function for a sample of kindergarten students in New York City. High 5s math clubs were designed to provide small group math enrichment delivered in a game-like format. Participants included 655 kindergarten students in 24 schools. Students assigned to the High 5s group met outside of class in small groups with a trained facilitator three times per week. The High 5s program produced a positive impact on one of two measures of math skills. We find no impact of the program on other outcomes.

Evaluating the Impact of Small Group Supplemental Math Enrichment in Kindergarten

Over the last few decades there has been a heavy emphasis on increasing literacy skills among low-income children, with federal and state initiatives designed to ensure that all children can read by grade 3. State and federal dollars have been spent to improve reading curricula, hire reading coaches, and to provide tutoring and small group support for struggling readers. However, much less emphasis has been placed on improving the early math skills of students in low-income schools. This is true despite the fact that research has shown that math skills are highly correlated not only with later math achievement but also with later reading achievement and other outcomes such as high school completion and college attendance (Jordan, Kaplan, Ramineni, & Locuniak, 2009; Duncan et al., 2007; Duncan & Magnuson, 2009). Focusing on early math skills, then, has the potential to improve child outcomes and reduce educational disparities. However, there is currently only limited information available regarding the effectiveness of early math instructional interventions on improving children's outcomes, especially in the early elementary school years. In particular, while there is substantial evidence of the effectiveness of small group instruction for reading, there is almost no information about the effectiveness of small group instruction in math.

This paper reports findings from a one-year randomized control trial investigating the effectiveness of a small group "math club" enrichment program for kindergarten students, called High 5s. Children from 24 New York City public schools participated in the program. All students who participated in the study had also been exposed to the *Building Blocks* mathematics curriculum in preschool. In this paper, we compare the outcomes of children who did and did not receive the High 5s program in kindergarten and address the question of whether or not participating in the High 5s program improved math and other achievement outcomes for the children who participated.

Background and Literature Review

As noted above, early math skills have been found to be highly predictive of later school success across a variety of outcomes. Not only do strong math skills predict later achievement in math but they are also predictive of later reading achievement and executive function skills (Claessens & Engel, 2013; Claessens, Duncan, & Engel, 2009; Duncan et al., 2007, Duncan & Magnuson, 2009). It has been hypothesized that this may be because to solve math problems students must manipulate numbers and shapes and they may also be expanding their vocabulary as they seek to explain their answers and justify their mathematical thinking (Ginsburg, Lee, & Boyd, 2008; Clements, Sarama, & Germeroth, 2016).

Unfortunately, kindergarten math instruction, in its current instantiation, may not provide opportunities to engage in the types of higher-order mathematical thinking that would lead to broad impacts. In effective mathematics classrooms, students are actively engaged in doing mathematics, solving challenging problems, making interdisciplinary connections, sharing mathematical ideas, using multiple representations to communicate mathematical ideas and using manipulatives and other tools (Protheroe, 2007; National Council on Teachers of Mathematics, 2000). However, research has shown that kindergarten math instruction is often very basic with most students entering kindergarten already possessing the skills that they are taught (e.g., basic counting, basic shape recognition) and thus likely does not provide the types of challenging activities that might elicit broader skill development. Indeed, additional time spent on these basic skills is associated with lower math achievement at the end of kindergarten while time spent on more advanced skills (e.g., addition and subtraction) is associated with higher math achievement (Engel, Claessens, & Finch, 2013; Engel, Claessens, Watts, & Farkas, 2016).).

In urban districts, with large class sizes and few extra adults available to help in the classroom, instruction is often limited to whole group instruction. Small group math instruction has the potential to provide a richer, more engaging math experience for young children. Research suggests that small group instruction provides more opportunities for individualized instruction, hands-on activities, and peer interaction and discussion that are more difficult to achieve in a large group setting (Webb, 1991; Yackel, Cobb, & Wood, 1991). While small groups are beneficial for all students, they are particularly developmentally appropriate for this age group. Young children are in a critical language development period and small groups provide opportunities for language acquisition and comprehension (Wasik, 2008; Phillips & Twardosz, 2003). Furthermore, positive teacher-child relationships and one-on-one interaction with adults are one of the strongest influences on learning at this age (Bowman, Donovan, & Burns, 2001).

Studies of literacy interventions have consistently shown that small group instruction has positive impacts on reading achievement (e.g. Slavin, Lake, Davis, & Madden, 2010). That holds true for studies where the groups were led by teachers and also those in which the groups were led by para-professional staff or volunteers (Slavin et al, 2010; Vadasy, Sanders, & Peyton, 2006; Vadasy & Sanders, 2008). Much less is known about benefits of small group math instruction. Several studies of one-on-one tutoring in math have demonstrated positive results. A meta-analysis by Ritter, Barnett, Denny, and Albin (2009) found among five studies of one-on-one tutoring programs in math (with sample sizes that ranged from 22 children to 385), the average effect size was +0.27. However, given the small number of studies, this average impact was not statistically significant at the .05 level. Fryer (2014) explored the impact of small group (three-on-one or two-on-one) tutoring for 4^{th}, 6^{th} and 9^{th} grade students. For the fourth graders, program impacts came from both experimental and quasi-experimental analyses. For the 6^{th} and 9^{th} graders, estimates were limited to quasi-experimental analyses. Tutors were recent college graduates and were paid a nominal salary and provided with room and board. Student who received small group tutoring outperformed their non-tutored peers in all three grades although the differences were only statistically significant for the 6^{th} and 9^{th} grade samples. Finally, Smith, Cobb, Farran, Cordray, and Munter (2013) investigated the impact of the Math Recovery program for first grade students who are struggling in math. The program provided one-on-one tutoring for 30 minutes four or five times a week for 12 weeks. The authors found positive and statistically significant impacts of the program on all six math outcomes they measured.

There have been few rigorous evaluations that have assessed the effectiveness of small group math instruction for young children. A handful of studies have compared different approaches to small group tutoring, but none have compared small group tutoring to a control group that was not receiving an alternative treatment, and almost all the studies that have explored the impact of small group work in math have targeted students who were at-risk for math difficulties. This study explores the impact of supplemental small group math instruction on kindergarten students more generally.

The High 5s Program

The High 5s program, a small group "math club" approach to supplemental math instruction, was developed at the University of Michigan with input from both MDRC and the developers of *Building Blocks* and is based on their research on mathematical learning trajectories (Clements & Sarama, 2014). The program was intended to provide a small-group math enrichment experience that was aligned with both the content and the approach of the *Building Blocks* curriculum to which the kindergarten students in the study had been exposed in pre-K. The *Building Blocks* program focuses on moving children along mathematical learning trajectories—the developmental progressions through which children learn mathematics. Each trajectory includes a mathematical goal, a developmental path along which children progress on the way to achieving that goal, and a set of instructional activities that can help children develop higher levels of thinking as they progress toward that goal. *Building Blocks* also encourages such instructional practices such as hands-on learning, student reflection about mathematical thinking, formative assessment to measure student progress and allow teachers to modify their approaches, instruction differentiated by children's ability levels, and a mix of small- and whole-group activities (Clements & Sarama, 2013). High 5s was designed to include these key elements as well.

The High 5s program was implemented by Bank Street College of Education, which hired, trained, and supervised the facilitators that ran the clubs. Clubs met three times a week for approximately 30 minutes and the curriculum included material for up to 28 weeks of instruction. In this study, clubs were held during non-instructional time (before school, after school, or during lunch) and ran from October 2015 through the first week of June 2016. Activities in the clubs are delivered in a game-like format and were intended to be fun, engaging, and developmentally appropriate. Most clubs included 4 children working with a facilitator. Facilitators were paid a salary comparable to that of a para-professional. A total of 79 clubs operated across the 24 schools participating in the study, with each school hosting approximately three clubs. Facilitators led three to four clubs each. Students were assigned to clubs based on logistical considerations (e.g. it was easier to schedule students for afterschool clubs if they were attending after-care at the school) and parents' scheduling preferences. Facilitators were assigned to clubs in an effort to minimize travel times between schools.

Each club session includes two start-up activities and a main activity. The start-ups are short (3-5 minute) activities meant to help students adjust to the club setting and reinforce key skills, such as counting. Start-ups are typically repeated each session for one week, and together the two start-ups are intended to last between 7 and 10 minutes. The main activities, designed to last 15 to 20 minutes, are the key component of clubs. The main activities are designed to deepen children's understanding of key concepts and to move them along the developmental trajectories. Each main activity is intended to be repeated up to three times over the course of the year, but not in consecutive sessions.

Each activity (both start-up and main activities) identifies a mathematical objective, indicates the mathematical development levels being targeted, and includes a semi-scripted activity plan. The activity plan gives facilitators enough support to implement the activities with fidelity but also the flexibility to include adaptations as necessary. Suggestions are also given for how to "scaffold" students at different levels of development (that is, how to provide assistance based on a student's current level of ability and extra support or additional challenges to help the child move to the next level) and draw out student thinking about the mathematics.

Every fourth club session is a Game Day. During Game Day, students are given the opportunity to choose from three or four activities set out by the facilitator. Facilitators can choose from new activities designed just for Game Day or from main activities that students have played previously in a regular club session. Game Days provide students with an opportunity to make choices, add variety, and allow students to interact with materials independently. They also provide facilitators with an opportunity to reinforce concepts on which students needed more practice and to work with students who needed individualized instruction. (See Jacob, Erickson & Mattera, 2018 for more detailed information about the High 5s program; a sample activity is included in the Appendix).

Club Facilitators

Twenty-four facilitators (most with bachelor's degrees but with limited formal teaching experience) ran the High 5s math clubs. Two facilitators took a maternity leave during the year and one left the project for personal reasons. To cover the clubs originally run by these facilitators, one full-time and two substitute facilitators were hired, so the number of facilitators at different time points varies. To be eligible for the position, facilitators had to have some prior experience working with children and at least an associate's degree. The facilitators were supported by five supervisors and a program director at Bank Street College. Bank Street also provided administrative, information technology, and human resources support for the program.

As shown in Table 1, the group of facilitators hired for the project was diverse. Facilitators ranged in age from 22 to 39 and came from a variety of racial and ethnic backgrounds. A majority were recent college graduates who had an interest in education and/or who had taken courses in education or related areas. They were mostly female (75 percent), and 29 percent identified as Hispanic or Latino, 38 percent as white, 29 percent as black, and 4 percent as another race. One-quarter were fluent in Spanish and 83 percent had a bachelor's degree. Just over a quarter of the facilitators had their highest degree in the field of education, but none had a New York State teaching certificate. They averaged less than two years of formal teaching experience, defined as an "assistant or lead teacher, including student teaching"; about one-third (29 percent) had no formal teaching experience. Most had some experience working with children, defined as experience in "a non-academic or non-classroom setting (e.g., tutoring, summer camp, after school program).” Facilitators were paid a salary commensurate with that of a paraprofessional teacher in the New York City public schools (around $25 per hour, depending on experience).

Facilitator Training

The High 5s model involved a substantial amount of training and supervision designed to support facilitators over the course of the year. Facilitators received 16 days of training before clubs began (about 6 of which were spent on research-related activities, such as learning to log information about the clubs for research purposes, or administrative tasks, like technology setup and fingerprinting). The rest of the training — conducted by staff members from Bank Street, MDRC, and researchers from the University of Michigan and the University of Denver— incorporated a variety of topics, including how to teach mathematics in an age-appropriate manner and how to facilitate small groups, as well as practice implementing the curricular activities. Facilitators were also trained on how children learn mathematics and how they develop mathematical skills. Finally, facilitators were given an opportunity to observe in a pre-K or kindergarten classroom.

All facilitators completed a certification process at the end of training. Facilitators were asked to conduct a short version of a club (one start-up and the main activity) from beginning to end. To be certified, facilitators had to meet the time requirements and receive a 3 or above (on a 5-point scale) on all the specified instructional quality and fidelity items.

The facilitators also received an additional eight days of training throughout the school year. These training sessions were held on days when the public schools were closed for in-service or school breaks. Topics of training sessions held during the year were chosen in response to the needs of the facilitators as expressed to supervisors or that came to light after reviewing daily logs, and included, for example, activities designed to help facilitators manage small groups, differentiate instruction, and facilitate mathematical reflection.

Finally, supervisors met with their cohort of four to five facilitators weekly. Cohort meetings included support with logistics, curriculum review, reflection about students, and professional development. Supervisors also met individually with facilitators regularly and provided coaching in the field as needed. Supervisors reviewed data from the daily facilitator logs to monitor timely completion, ensure that activities were conducted as scheduled, and to identify any curricular challenges, behavioral or attendance issues, or logistical problems.

Theory of Change

As depicted in Figure 1, there were a number of components of the High 5s program that may have had an impact on student achievement. First, since the clubs provided supplemental instruction for students that took place outside of regular instructional time, students might simply benefit from additional time spent on math activities. Prior research on the impact of additional instructional time on student learning has been somewhat inconclusive (Hanushek, 2015). Although many studies have demonstrated a link between increased instructional time or time on task and achievement (e.g. Lavy, 2015; Rivkin & Shiman, 2015; Marcotte & Hemelt, 2008), not all programs that add additional instructional time lead to increases in achievement (e.g. Gamse, Bloom, Kemple & Jacob, 2008). Further, research shows that the effects of additional time vary based on a variety of factors including the quality of the instruction and the classroom environment (Stallings, 1980; Lavy, 2015; Rivkin & Shiman, 2015; Cattaneo, Ogenfuss, & Walter, 2016).

Second, because the clubs involved instruction in small groups, we hypothesized that students might receive more individualized and differentiated instruction. Small groups make it easier for instructional leaders to know what each student understands and makes it harder for students to "hide" or become disengaged (e.g. Webb, 1991; Yackel et al., 1991; Wasik, 2008). At the same time, Wasik (2008) notes that small groups can encourage more positive interactions between children and their teachers. Research supports the notion that the more quality contact the young child has with a competent adult, the greater the positive impact it can have on learning and development (Darling-Hammond & Bransford, 2006; Pianta, 2006). The small group nature of the High 5s clubs was designed to promote more quality time with an adult and the curriculum was designed to encourage differentiation, with specific suggestions for altering activities for students at different skill levels.

Third, because the High 5s clubs were designed to pick up where the Building Blocks curriculum left off, the curricular materials assumed a somewhat higher level of mathematical knowledge than what the typical kindergarten student might enter kindergarten with. Thus, we hypothesized that students in High 5s might be exposed to more advanced mathematical content than they would be in their classrooms. Research shows that exposing kindergarteners to more advanced mathematical material can lead to improvements in academic achievement (Engel et al., 2016).

Fourth, High 5s clubs all involved hands-on activities in which students are actively engaged with mathematical materials, and we hypothesized that this type of hands-on learning might lead to improved outcomes. Theory on child development underscores the importance of play based approaches to learning for young children (Bodrova & Leong, 1996; Bredekamp, 1987; Bredekamp & Rosegrant, 1992) and research documents the benefits of using manipulatives in mathematics instruction (Carbonneau, Marley, & Selig, 2013). Similarly, we hypothesized that the playful nature of the activities would lead to increased engagement and a more positive attitude towards mathematics. The relationship between children's attitudes toward math and math ability is reciprocal, so positive feelings towards math can help develop math ability (Fisher et al., 2012).

Finally, we hoped that the training and small group format would help encourage facilitators to engage in mathematical reflection with students, by underscoring the mathematical objective of each activity, asking open-ended questions (i.e. questions with more than one right answer) and summarizing what was learned. Best practices in mathematics instruction argue for the importance of helping students see mathematical connections, and talk about mathematics and their mathematical reasoning, as a way to develop deeper understanding and to help students see the utility of mathematics (National Council of Teachers of Mathematics, 2000).

Methods

The study was designed to address the following research questions:

Research Question 1: Can a small group math program be delivered with quality and fidelity by paraprofessional-level staff?

Research Question 2: How does the instruction in the High 5s program compare to the math instruction being delivered in kindergarten classrooms?

Research Question 3: Among students who were exposed to *Building Blocks* in preschool, what is the impact of the High 5s kindergarten program on math skills, attitudes toward math, executive function and verbal skills at the end of kindergarten?

Sample

The High 5s kindergarten study took place in the 24 public schools that had implemented the *Building Blocks* program in pre-K as part of the Making Pre-K Count study (see Mattera, Jacob & Morris, 2018 for details). All 24 schools primarily served students from low-income families. Recruitment of schools was done by contacting principals/assistant principals of the public schools participated in MPC who then signed an MOU agreeing to participate in the High 5s kindergarten study and receive the High 5s program for a subset of their kindergarten students. All eligible schools (n = 24) chose to participate. Schools received the High 5s program free of charge.

Kindergarten students were eligible to participate in the High 5s program if they were enrolled in one of the 24 participating schools and had attended pre-K in that same school the previous year. In the spring of pre-K, 97% of eligible families gave consent for their children to participate in the High 5s study. In the fall of the kindergarten year, individual children were randomly assigned within schools to either receive the High 5s program in addition to their regular kindergarten math instruction (n=320) or to a "kindergarten-as-usual" control group (n=335). Some students who were assigned to the High 5s program group did not participate in the program or discontinued their participation before the end of the year. Of the 303 who assigned to High 5s and were located for assessment at the end of the year, 284 were still participating in the program at the end of the year. All students who were assessed at the end of kindergarten, regardless of their participation, were included in analyses.[1,2]

---

[1] The analytic sample includes 18 students whose parents did not give consent to participate in High 5s, but who were randomized in a second phase to maintain the internal validity of the comparison with a control group of children who did not receive Making Pre-K Count in pre-K. Students in the second randomization phase did not receive the High 5s program even if they were assigned to the treatment group. The analytic sample also only includes those students who remained in the same school between preK and K. Results without these 18 students and including students who did not remain in the same school between preK and K are the same as those presented here. See Mattera et al. (2018) for details.

[2] While all students, regardless of participation, were included in analyses of child outcomes, the 19 students who discontinued participation in the program were dropped from attendance totals once they were no longer participating (e.g. after they had left the school).

As shown in Table 2, there were no statistically significant differences in measured baseline demographic characteristics or baseline skills between children who received High 5s in addition to their kindergarten instruction and the kindergarten as usual group.

Kindergarten classrooms contained students from both program and control groups as well as other students who were not part of the *Building Blocks* or High 5s studies. The average classroom had 20 students and approximately 3 students had received both *Building Blocks* in pre-K and High 5s in kindergarten and around 3 had received *Building Blocks* in pre-K but did not receive High 5s in kindergarten. The remaining students in these classrooms were not part of the *Building Blocks* or High 5s studies. To reduce the likelihood of contamination, High 5s clubs were held in pre-K classrooms or other multi-use spaces (e.g., a resource room), and not in kindergarten classrooms. In an effort to keep teachers informed, kindergarten teachers were provided with a few example activities, but otherwise teachers did not have access to the High 5s curriculum. Most of the interactions that High 5s facilitators had with teachers were brief, and mostly focused on dropping off or picking up students from the classroom. While it is possible that there was some contamination at the student level, any such contamination would bias our impacts towards zero.

Attrition Analysis

We were able to obtain student achievement data from 622 of the 655 students who were randomly assigned as part of the study (Figure 2). This yields an overall attrition rate of 5%. Attrition rates were similar for the treatment and control groups with a differential attrition rate of 2%. These rates are unlikely to lead to bias in our results (What Works Clearinghouse, 2017). Moving out of the New York City area and parent or child refusal were the two primary reasons for students not being assessed.

Measures of Implementation

The strength of implementation was assessed using (1) daily logs completed by facilitators and (2) via club observations conducted by study staff. The primary method for measuring the structural aspects of implementation fidelity were the daily logs. Facilitators completed logs each day, for each of their clubs, in an online management information system (MIS). This resulted in a total of more than 7,000 logs collected over the course of the year. Research has demonstrated that logs are a reliable, valid and cost effective way to measure a number of aspects of instructional practice (Rowan, Camburn, & Correnti, 2004)

The daily logs included information on a) child attendance, b) the activities that were conducted that day, c) the length of time spent on each activity, d) the level of child engagement with each of the activities, e) the degree to which child behavior was an issue in the clubs, f) the degree to which facilitators experienced any logistical or other problems with the clubs. (A sample log is included in the Appendix). As noted above, supervisors had access to the logs and used the information provided to identify facilitators and/or specific clubs that were in need of additional support—typically those in which the facilitator reported consistently low levels of student engagement or ongoing instances of student behavior problems. Supervisors also used the log data to identify children with low levels of attendance and followed up with individual families to try to eliminate barriers to attendance.

In addition to the daily logs, direct observation of clubs were conducted in both the fall of 2015 and the spring of 2016. Observations were conducted with the goal of verifying the log data and gathering additional information about instructional quality. Fall observations took place in November. Spring observations were conducted in late February and March. Each facilitator (N=24) was observed at least once in each time period and at least one observation was conducted in every High 5s school (N=24). In order to observe in all schools and to observe each of the facilitators, it was necessary to conduct more than 24 observations at each time point. One facilitator was observed twice in the fall and spring and in the spring both substitute facilitators were observed. When possible, each facilitator was observed in the same club in both the fall and spring.

Club observations were conducted using a researcher-developed instrument (See the Appendix for a copy of the observation protocol). The instrument was designed to align, as much as possible, with the observational measure used in the kindergarten classroom observations, described below. Observers were members of the research team and outside consultants who had received training and been certified on the instrument. The instrument included items related to adherence to the curriculum, barriers to implementation, instructional quality, classroom management, and student engagement. Most items were rated on a 5-point scale (1=rarely or never to 5=always or almost always). In a few cases, items were dichotomous (yes or no), or observers hand-entered information (e.g., start and end times).

In this study we distinguish between what we will call structural fidelity (adherence to program design including conducting the activities as written on the correct days, spending sufficient time on math, etc.) and process fidelity (quality of instruction, instructional climate, etc.) (Century, Rudnick & Freeman, 2010; Mowbray, Holter, Teague, & Bybee, 2003; O'Donnell, 2008). Observations were intended to provide additional information on which to assess fidelity of implementation and in particular to assess the process aspects of fidelity. Our observations were conducted twice, to mirror the protocol used in other recent studies of early childhood educational interventions (Morris, et. al., 2014; Clements & Sarama, 2008; Clements & Sarama., 2011). In all these studies researchers were balancing the desire to collect in-depth, high quality information on instructional quality (which requires substantial training and field time), and cost. Combined with the daily logs, these two sources of information provide a rich set of data with which to assess fidelity.

Measures of Program Contrast

All students who were part of the High 5s study also received typical math instruction in their kindergarten classrooms. Because the High 5s program is supplemental and took place outside of regular instructional time, students in the High 5s program were spending more time on math each week than children in the kindergarten as usual control condition. However, as shown in Figure 1, we also hypothesized that the program could impact math achievement via other pathways as well, including offering more opportunities for differentiated instruction, increased exposure to more advanced mathematical content, more hands-on learning experiences and additional opportunities for mathematical reflection. To test these hypotheses, and to better understand the potential value-added of the High 5s program the research team observed the mathematics instruction that children were receiving in their classrooms.

The team observed kindergarten classrooms in the 24 High 5s schools in the fall and spring of the kindergarten year. A total of 75 observations were conducted in 42 participating classrooms. When a teacher who was observed in the fall was not available in the spring, a replacement teacher was selected for observation. Thirty-three teachers were observed in both the fall and the spring. Three teachers were observed only in the fall, and six new teachers were observed in the spring. As noted above, the decision to observe classrooms twice was made to mirror protocols used in other similar studies of early childhood settings (Morris, et. al., 2014; Clements & Sarama, 2008; Clements & Sarama, 2011) and to balance the quality and cost of the observation. Observing in both the fall and the spring is intended to capture changes in instructional practice and content coverage that occur as the school year progresses and to allow for more than a single snapshot of classroom instructional climate.

The observation protocol included a simplified version of the Narrative Record and an adapted version of the Classroom Observation of Early Mathematics — Environment and Teaching (COEMET) (Sarama & Clements, 2007). The Narrative Record (Farran & Bilbrey, 2004) is an open-ended format for describing the types of activities, the content of instruction and the amount of time devoted to each activity. The COEMET describes math instruction during formal math activities, and items related to math practices, type of activity, student engagement, and instructional quality. As with the club observations, kindergarten classroom observers received training and were certified on the observation instruments before they were able to conduct observations in the field. Several of the items in the observation protocol were identical to items that were included in the High 5s club observation protocol, which helped facilitate comparisons. Approximately 20 percent of the observations were double coded to ensure reliability. Cohen's kappa for the double-coded observations was 0.92.

In addition, to help strengthen our inferences, the data collected during classroom observations were compared to the information provided in the GoMath! curriculum used in most of the kindergarten classrooms. Teachers in the classrooms we observed adhered closely to the curriculum.

Student Outcome Measures

To measure student achievement, direct child assessments were administered in spring 2016, at the end of kindergarten. Child assessments were conducted between March 22 and June 29, 2016. Assessments were conducted by a survey firm who went to individual schools to conduct the assessments. Assessors received training, were certified to conduct assessments, and were blind to intervention status. Students within the same school were assessed during the same short window of time (usually a few days). There were no statistically significant differences in the timing of assessments between the treatment and control groups. Student outcomes included math skills, math attitudes, language skills, and executive function skills. Math skills were assessed using two measures; the Research-Based Early Math Assessment – Kindergarten (REMA-K, Clements, Sarama, & Liu, 2008) and the Woodcock-Johnson Applied Problems subscale of the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, & Mather, 2001).

The REMA-K is a direct assessment that measures thinking and learning along research-based developmental progressions for math topics. To shorten the total duration of each assessment, this study used an adaptation of the Research-Based Early Math Assessment. The items selected for the adapted version represents the full range of early mathematics competencies applicable within the prekindergarten, kindergarten, and first-grade years.

The Woodcock-Johnson Applied Problems is a subscale of the Woodcock-Johnson III Tests of Achievement. It is a valid standardized assessment of mathematical thinking for ages 2 through 90; items are suitable for assessing simple math functions relevant at young ages (such as identifying the number when more objects are added to a picture). The Woodcock-Johnson Applied Problems test is a less detailed, more global measure of children's math ability than the REMA-K.

The research team also created an assessment to measure children's attitudes toward math. The assessment, which asked children to indicate, by pointing to a series of five faces (sad to smiling) how happy or unhappy school and math made them feel, was based on a set of items that had been used previously to assess children's attitudes toward math (White, 2015; Thomas

& Dowker, 2000; see the Appendix for a copy of the instrument). A rating of 1 indicated that they felt very unhappy and a rating of 5 indicated that they felt very happy about math. This item was also used to assess the math attitudes of students who participated in the enhanced preschool program that was evaluated as part of the larger study in which the High 5s evaluation was embedded. In that study, the enhanced preschool program was found to have a positive impact on students' attitudes toward math, suggesting that the item(s) were sensitive enough to capture differences (Mattera et al., 2018).

Language skills were assessed using the Receptive One-Word Picture Vocabulary Test (ROWPVT-4; Martin & Brownell, 2011). The Receptive One-Word Picture Vocabulary Test (ROWPVT-4) assesses children's receptive vocabulary, or their ability to understand spoken language, by asking them to match a word the assessor says out loud to a picture of an object, an action, or a concept.

Executive function was measured using the Hearts and Flowers (Wright & Diamond, 2014) computerized task; and the Corsi Blocks task (Corsi, 1972; Lezak, 1983). Hearts and Flowers is a computerized task that measures inhibitory control. During this task, a child is asked to select the button on the same side of the screen if a heart appears on the screen and on the opposite side of the screen if a flower appears on the screen.

The Corsi Blocks task assesses short-term working memory. During this task, a child is asked to repeat a sequence of blocks tapped by an assessor, tapping the blocks in reverse order. The child begins with a sequence of two blocks and more blocks are added to the sequence.

Analysis

Program impacts were estimated by comparing the mean outcomes for children assigned to High 5s as a supplement to their kindergarten math instruction and to children assigned to the kindergarten as usual control group. The models control for selected background characteristics and include a dummy variable for each school site. Random assignment for High 5s took place in the fall of the kindergarten year, therefore, baseline covariates for children come from the spring of their pre-K year. In this study design, treatment group students are nested within clubs, and all students were nested within both classrooms and schools. Either fixed effects or random effects can be used to address this clustering, although neither approach perfectly addresses the problem (Weiss, Lockwood & McCaffery, 2014). We have chosen to use a fixed effects approach because we were intending to perform finite sample inference that conditions on our sample and the post-randomization groups (Weiss, Lockwood & McCaffery, 2014). However, we have also conducted robustness checks in which we run a generalized model that includes cluster-adjusted standard errors, which is akin to random effects, and which accounts for any clustering of students that may occur within schools (see for example, Visher et al., 2012). The standard errors and point estimates on these model vary slightly from what is reported here, but the resulting inferences remain identical. Alternative models are available from the authors upon request.

The following single-level model for child outcomes is used:

$$Y_s = \alpha_0 + \sum_{i=1}^{10} \alpha_i X_{is} + \pi T_s + \sum_{c=1}^{24} \alpha_c Z_{cs} + \varepsilon_s$$

where:

Y_s = the outcome for student *s*

X_{is} = baseline characteristic *i* for student *s*

Z_{cs} = an indicator variable for school *c* for student *s*

T_s = the treatment indicator, which equals one if student *s* was randomized to treatment (High 5s) and zero if the student was randomized to control status

ε_s = a random error for student s that is independently and identically distributed across students in classrooms

A subset of background characteristics were selected as covariates based on their degree of correlation with the outcome of interest and theoretical importance. Missing covariates, but not outcome data, were imputed using multiple imputation based on other available covariates and baseline assessments. Models included the following covariates:

- Whether the parent had a high school diploma/GED or a higher degree
- The child's age at the time of kindergarten spring assessment
- A measure of the child's level of English proficiency in the fall of pre-K (assessed by the pre-LAS)
- Three measures of baseline executive function assessing attention, inhibition, cognitive flexibility, and working memory (as measured by the Spatial Conflict Arrows task, the Pencil Tap task, and Corsi Blocks Forward Span)
- An evaluation by the examiner of the child's attention and impulse control during assessment administration at baseline (PSRA: Attention-Inhibition)
- A measure of baseline receptive language skills (ROWPVT)
- Two measures of the child's baseline math competency (ECLS-B scale score and Woodcock Johnson raw score)

The study team did not explicitly adjust p-values to correct for multiple comparisons (Bloom, 2004; Shadish, Cook & Campbell, 2003). Instead, we limited the number of outcomes in our main impact analysis and consider all subgroup analyses as exploratory.

Results

In this section we describe the implementation of the High 5s program, the degree to which it differed from regular kindergarten mathematics instruction, and discuss the impacts of the program on math skills, attitudes towards math, language ability, and executive function.

Program Implementation

We begin by assessing how well the program was implemented. We used both structural (e.g., facilitators conduct activities as scheduled) and process (e.g., facilitators ask open ended questions to elicit student responses) measures to determine if the High 5s program was delivered as intended.

As shown in Table 3 findings indicated that the structural aspects of implementation were strong. Student attendance was high. Average overall attendance was over 87 percent. Facilitator logs indicate that clubs were conducted as scheduled. Ninety-three percent of scheduled sessions were completed. Sessions were canceled primarily due to changes in school schedules, unavailability of space, or an insufficient number of students in attendance (clubs were only held if at least two students were in attendance). Facilitators also followed the intended pacing of the curriculum and activities were completed as scheduled (i.e. the correct activities were conducted on the appropriate day). According to club logs, facilitators conducted the specific activities that were scheduled for that day in 96 percent of clubs. Observations also indicated that activities were typically implemented as intended with few mathematical errors. As shown in Table 4, on average, students were engaged in almost 25 minutes of math each session, exceeding the benchmark of 20 minutes of math per session that had been established at the beginning of the program.

Instructional quality, or process implementation, was more varied. Table 5 shows the average ratings given by observers in both fall and spring observations. Each measure was rated on a 5-point scale (1-5), where 3 was considered satisfactory and 4 exceeded expectations. While overall ratings were typically above the satisfactory level, facilitators were rated more highly on items related to creating a positive instructional climate in the clubs, but had more difficulty consistently asking open-ended questions or engaging in mathematical reflection. Despite additional training on these topics, there was little improvement observed between the fall and the spring. The average ratings also mask some of the underlying variability among facilitators. To capture this variability, we created a composite score for each facilitator, averaging across all measures of instructional quality given by observers in the fall and the spring. Among all the facilitators, 27% had composite scores of 4 or above, 54% had scores between 3 and 4, and 19% had scores below 3.

Comparison to the Kindergarten Environment

Classroom observations were designed to assess the degree to which the instruction received in High 5s differed from the instruction provided in classrooms. We assessed how much additional time in math instruction students in the High 5s program received compared to students who only received instruction in class. The study team also explored the types of mathematical materials and instructional groupings to which students were exposed in their kindergarten classrooms as well as the degree to which teachers and facilitators asked open-ended questions, differentiated instruction and encouraged mathematical reflection. Finally, the team compared the mathematical content that was covered in classrooms and in High 5s clubs. Table 6 summarizes the findings. Because the data collection instruments differed, not all items are directly comparable, and thus we did not conduct statistical tests of the differences.

However, they are included because they provide insight into potential differences in the nature of instruction.

Time. Kindergarten teachers spent an average of 56 minutes on math each day, or about 280 minutes of math each week. High 5s added an additional 75 minutes each week, which is an increase of 27 percent.

Grouping and Materials. Most of the instructional time in classrooms was spent in whole group instruction or doing seatwork, in which students worked individually at their seats while the teacher rotated around the classroom. Students spent approximately 5 percent of their mathematics time working in small groups (primarily in pairs). During kindergarten math instruction, students spent about 35% of their time engaged in activities that used some sort of mathematical materials (other than workbooks or worksheets). By design, each High 5s session was conducted in a small group and all activities involved hands-on materials.

Mathematical Reflection and Differentiation. There were also some differences in the degree of differentiation and mathematical reflection between classroom instruction and the instruction received in High 5s. On average, High 5s facilitators were rated somewhat higher than classroom teachers on scales measuring these aspects of instructional quality: asking open-ended questions (+0.7 points), encouraging mathematical reflection (+0.5 points), and scaffolding to extend math skills (+0.4 points) (Table 6), although ratings were relatively low for both groups.[3] Facilitators also differentiated their instruction in almost 58 percent of observations, versus 36 percent for teachers.

Content. Table 7 shows a comparison between the mathematics content covered in the High 5s clubs and in the kindergarten classrooms. Information on kindergarten content was

[3] Comparisons were only possible in the spring. Items were not aligned well enough on the fall protocols.

measured during classroom observations at two time points (fall and spring) and verified by comparing observed content to the written curriculum (GoMath!) used in most kindergarten classrooms. Findings indicate that High 5s clubs spent less time on basic counting (i.e., counting forward by ones) and more time on more advanced counting skills (i.e., counting backward by 1s, counting forward or backward by 10s, skip counting by a number other than 10, and counting forward by 10s and 1s) than kindergarten classrooms. For example, students in kindergarten classrooms spent over 20% of their time in both the fall and spring counting forward by 1s, while students spent no time counting forward by 1s in High 5s clubs. On the other hand, over 20 percent of the time in High 5s clubs was spent engaged with complex counting compared to approximately 5 percent of the time spent on this type of counting in kindergarten classrooms. In kindergarten classrooms students spent slightly more time on addition and subtraction than High 5s clubs did, while High 5s clubs spent significantly more time on other math content (e.g., patterning, shapes, and measurement- 21 percent in High 5s vs. less than 10 percent in kindergarten classrooms). Chapters on geometry and measurement come at the end of the district-supported Go Math! curriculum, and it is possible that teachers may have covered these topics later in the year. However, preliminary analyses of pacing in these classrooms suggests that most teachers would not have reached these chapters by the end of the year.

Impacts of the Program on Student Achievement

The addition of the High 5s kindergarten program over and above regular kindergarten math instruction produced a positive impact on children’s REMA-K scores (Table 8; effect size = 0.20). Children who received the High 5s kindergarten supplement had an average score of 39.55 on the REMA-K, and children who received the kindergarten curriculum had a score of 37.83 on the same assessment. This difference is equivalent to approximately two and a half months of additional growth on the REMA-K. This effect is equivalent to closing almost one-fifth of the achievement gap between low-income children and their higher-income peers.[4]

As shown in Table 8, there was a positive but not statistically significant impact on the Woodcock-Johnson Applied Problems test (effect size = 0.09). The High 5s program did not have an effect on children's math attitudes, language, and executive function outcomes. Subgroup analyses were conducted for six child level characteristics:

- Baseline math skills
- Baseline inibitory control
- Baseline receptive language
- gender
- Spanish-speaking ability

There were very few statistically signficant differences between subgroups and no consistent pattern of findings (See Mattera et al., 2018 for details). For example, as shown in Table 9, High 5s had a statistically significant impact on the Woodcock-Johnson scores of children who had high inhibitory control skills at kindergarten entry (effect size = 0.17) but not for children who entered kindergarten with low inhibitory control.

Although the subgroup findings suggest that the High 5s program may have had an impact on outcomes other than the REMA-K for some subgroups of children, the findings should be interpreted with caution. There is no consistent group of children that benefited across all

[4] The standardized measures of the difference in outcomes at the end of kindergarten for children in the 90th income percentile and children in the 10th income percentile in the Early Childhood Longitudinal Study, Kindergarten Class (ECLS-K) conducted in the 2010-2011 year (as described in Reardon & Portilla, 2016) was equivalent to 1.046 standardized units. The estimated effect size produced by the High 5s clubs (0.20) is equivalent to 19 percent of that gap.

outcomes, most of the differences were not statistically different from one another and would not remain statistically significant after accounting for multiple hypothesis testing.

Finally, to explore how much the additional instructional time that the High 5s program added mattered, we conducted a subgroup analysis to assess whether the impact of High 5s was smaller in schools in which students received more in-class math instruction. We hypothesized that additional time spent in High 5s might matter less in schools in which students were already receiving a substantial amount of math instruction (i.e., in schools where High 5s was adding a smaller percentage of additional time). Schools were divided into two groups—those that spent fewer than 53 minutes per day on math and those that spent more than 53 minutes per day on math, on average, based on classroom observations. Fifty-three minutes was selected because it represents the 50^{th} percentile for time spent on math in the classroom.

As shown in Table 10, the results were not consistent with the hypothesis. There was a statistically significant impact on the REMA-K (effect size= 0.29) among students in schools that spent more time per day on math. The impact was not statistically significant (effect size=0.10) for students in schools in which less time was spent on math in class. These effects were not statistically signficant from one another. Interestingly, High 5s also had a larger impact on children's math attitudes in schools with lower than average time spent on math (effect size = 0.29) than in schools with more time spent on math (effect size = -0.18), suggesting that the more time children spent in High 5s relative to classroom math instruction, the more likely they were to report positive attitudes toward math.

Discussion

The findings reported here suggest a number of potential mechanisms through which High 5s may have contributed to the development of children's math skills. First, High 5s led to additional mathematics instructional time. Students in High 5s received, on average, 75 minutes of math instruction each week over and above the approximately 280 minutes they were receiving in class. This is an increase of 27 percent, on average. However, High 5s was not more effective in schools that spent less time on math in the classrooms, compared to schools in which teachers spent more time on math in the classroom.

High 5s also took a different instructional approach than the typical kindergarten classroom. Students in the High 5s clubs experienced greater opportunities for differentiated and individualized instruction and more hands on learning opportunities. In classrooms, 85 percent of math time was spent in either whole-group instruction or seat work, and most activities involved either workbooks or no materials at all. High 5s instruction occurred in small groups or individually with the facilitator (on Game Days, some facilitators worked one on one with individual students). High 5s was primarily game-based and provided students with opportunities to engage with mathematical concepts using a variety of manipulatives and materials. These instructional practices were highly aligned with the Building Blocks curriculum, and this consistency may have contributed to student learning in High 5s.

High 5s also exposed students to more advanced mathematical topics. High 5s clubs covered a wider range of content than did kindergarten classrooms and more time was spent on more advanced mathematical skills.

Finally, the instructional climate in clubs may have differed from the climate in class-rooms. The following exchange, as reported by one of the facilitators during the last week of the clubs, captures this difference:

> While filling out their “why we like math” page, the children [in this club] all concluded that they didn’t like math. I [the facilitator] said that was strange because they’ve been

> doing math in High 5s all year and were loving it and were so happy. They clarified that they like math in High 5s but they don't like it in school. R explained in his words that "In school, you do math and you be quiet and look down at your paper. They just tell you that you're wrong. And then nobody talks to you. It's just wrong and you have to be quiet. But in High 5s we have you. You never say we did it wrong and we all talk and figure it out and then nobody's wrong. That's why I'm happy when I do math in High 5s."

Another facilitator wrote at the end of the year that the most rewarding aspect of the job was "When the kids finally 'get' a new concept and they become more confident in themselves it honestly makes it all worthwhile. One of my students wrote in her high 5s book that one of the things she learned in high 5s was to 'never stop trying.'" The extensive training and ongoing supervision that facilitators received from Bank Street likely contributed to this instructional climate. Given this, it is surprising that the study did not find a significant effect on children's attitudes towards math. As highlighted in the quote above, as well as the subgroup finding that students who spent less time engaging in math in the classroom had more positive attitudes toward math, the reason for this may be that when children answered survey items about their views of math they were thinking about their in-class math instruction and not about High 5s.

There is some evidence to support each of these potential contributions, and any effects of the program likely arise from some combination of all of them. It is therefore important to be cautious in concluding that just one or two of these elements would be sufficient to produce impacts.

Although this study provides some preliminary evidence that small group instruction may be a promising way to approach early math instruction, more research is required to replicate these findings and identify which elements are critical to achieving positive outcomes. For example, more work is needed to determine if it was simply the addition of extra time on mathematics, or the benefits of small group instruction or some combination thereof that contributed to the findings. Similarly, more work is needed to determine whether the program could be effective with less training and support provided to facilitators. In addition, the curriculum itself was designed to align with and build on what children in the study had experienced in pre-K. The study was not designed to demonstrate how effective the curriculum would be for children who did not have this grounding in preschool. Further research is needed on how these activities and this instructional setup would work for children who may have had a different set of experiences in the year before kindergarten. Finally, it will be important to understand whether the program's impacts can be sustained over the longer run. The study team plans to follow the children in the study through elementary school to understand the long term impacts of the program.

Exploring models that are less resource intensive will also be important. Although the program was delivered by facilitators who were paid a salary commensurate with that of paraprofessional staff, because the program operated outside of regular instructional time facilitators could each run only a few clubs, and often had to travel across the city to different schools. This structure also required substantial operational support from the research team to ensure that student attendance was high and space was available to meet in the schools each day. Other models, including conducting small groups in the regular classroom or running a pull-out program with existing school paraprofessionals, are being explored to see whether or not similar results could be obtained with fewer resources.

References

Bowman, B. T., Donovan, M. S., & Burns, M. S. (Eds.) (2001). *Eager to learn: Educating our preschoolers.* Washington, DC: National Academy Press.

Bodrova, E. & Leong, D.J. (1996). *Tools of the mind: The Vygotskian approach to early childhood education*. Columbus, OH: Prentice Hall/Merrill.

Bredekamp, S. (Ed.). (1987). *Developmentally appropriate practice in early childhood programs serving children from birth through age eight*. Washington, DC: Natl. Assn. for the Education of Young Children.

Bredekamp, S. & Rosegrant, T. (Eds.). (1992). R*eaching potential: Appropriate curriculum and assessment for young children*. Washington, DC: Natl. Assn. for the Education of Young Children.

Cattaneo, M.A., Oggenfuss, C. & Wolter, S.C. (2016). *The more, the better? The impact of instructional time on student performance* (Economics of Education Working Paper Series 0115). University of Zurich, Department of Business Administration (IBW).

Carbonneau, K. J., Marley, S., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*(2), 380-400. https://doi.org/10.1037/a0031084

Century, J.; Rudnick, M.; Freeman, C. (2010). A framework for measuring fidelity of implementation. *American Journal of Evaluation, 31*(2) 199-218 2010.

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review*, *28*(4), 415-427.

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*(6), 1-29.

Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science*, 333, 968– 970. doi:10.1126/science.1204537

Clements, D., & Sarama, J. (2013). *Building Blocks: Teacher's Edition.* Columbus, OH: McGraw-Hill.

Clements, D., & Sarama, J. (2014). *Learning and Teaching Early Math: The Learning Trajectories Approach*. New York: Routledge.

Clements, D., Sarama, J., & Germeroth, (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, *36*. 79-90.

Clements, D., Sarama, J., & Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Math Assessment. *Educational Psychology*, *28*(4). 457-482.

Clements, D. & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, 45, 443-494.

Corsi, P.M. (1972). *Human memory and the medial temporal region of the brain* (Doctoral dissertation). Montreal: McGill University.

Darling-Hammond, L., & Bransford, J. (2006). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey Bass.

Duncan, G., Dowsett, C., Claessens, A., Magnuson, K., Huston, A., Klebanov, P., Pagani, L., Feinstein, L., Engel, M., & Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6). 1428-1446.

Duncan, G., & Magnuson, K. (2009). *The nature and impact of early skills, attention, and behavior*. Paper presented at the Russell Sage Foundation Conference on Social Inequality and Educational Outcomes, New York City.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (Mis) Alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35*(2), 157-178.

Engel, M., Claessens, A., Watts, T., & Farkas, G. (2016). Mathematics content coverage and student learning in kindergarten. *Educational Researcher, 45*(5). 293-300.

Farran, D., & Bilbrey, C. (2004). *Narrative Record*. Unpublished instrument available from Dale C. Farran, Peabody Research Institute, Vanderbilt University, Nashville, TN.

Fisher, P. H., Dobbs-Oates, J., Doctoroff, G. L., & Arnold, D. H. (2012). Early math interest and the development of math skills. *Journal of Educational Psychology, 104*(3), 673-681. doi: 10.1037/a0027756

Fryer, R. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics, 129*(3), 1355-1407.

Gamse, B., Bloom, H., Kemple, J., & Jacob, R. (2008). *Reading First Impact Study: Interim Report*, US Department of Education: Institute for Education Sciences.

Ginsburg, H., Lee, J., & Boyd, J. (2008). *Mathematics education for young children: What it is and how to promote it*. (Social Policy Report 22), 1-24. Ann Arbor, MI: Society for Research in Child Development.

Hanushek, E.A. (2015). Time in education: introduction. *The Economic Journal*, *125* (588), F394-396.

Jacob, R., Erickson, A. & Mattera, S. (2018). *Launching Kindergarten Math Clubs: The Implementation of High 5s in New York City*, MDRC.

Jordan N., Kaplan D., Ramineni C., Locuniak M. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850–867.

Lavy, Victor. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *The Economic Journal*, *125*(588). F397-F424.

Lezak, M. (1983). *Neuropsychological Assessment*. New York: Oxford University Press.

Marcotte, D. E.; Hemelt, S. W. (2008). Unscheduled school closings and student performance. *Education Finance and Policy*, *3*(3), 316-338.

Mattera, S, Jacob, R. & Morris, P. (2018). *Strengthening Children's Math Skills with Enhanced Instruction*, MDRC.

Martin, N., & Brownell, R. (2011). *Receptive One-Word Picture Vocabulary Test*, 4th ed. Novato, CA: Academic Therapy Publications.

Morris, P., Mattera, S.K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to Improving Preschoolers' Social and Emotional Competence* (OPRE Report 2014-44). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*(3), 315–340. doi: 10.1177/109821400302400303

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics. Retrieved from https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*(1), 33–84. doi: 10.3102/0034654307313793

Pianta, R. (2006). *Teacher-child relationships and early literacy*. In: D. K. Dickinson & S. B. Neuman (Eds.), Handbook of early literacy research: Vol. II (pp. 149–162). New York: The Guilford Press.

Phillips, L. B., & Twardosz, S. (2003). Group size and storybook reading: Two-year-old children's verbal and nonverbal participation with books. *Early Education and Development, 14*. 453-478.

Protheroe, N. (2007). What does good math instruction look like? *Principal, 7*(1). 51-56.

Reardon, S., & Portilla, X. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open 2*(3).

Ritter, G., Barnett, J., Denny, G., & Albin, G. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research, 79*(1), 3-38.

Rivkin, S. & Shicman, J. C. (2015). Instructional time, classroom quality and academic achievement, *The Economic Journal, 125*(November), F425–F448. doi: 10.1111/ecoj.12315

Rowan, B., Camburn, E., & Correnti, R. (2004). Using Teacher Logs to Measure the Enacted Curriculum: A Study of Literacy Teaching in Third-Grade Classrooms. *The Elementary School Journal, 105*(1), 75-101. doi:10.1086/428803

Sarama, J., & Clements, D. (2007). *Manal for Classroom Observation of Early Mathematics: Environment and Teaching (COEMET) Version 3*. Unpublished manual.

Slavin, R., Lake, C., Davis, S., & Madden, N. (2010). *Identifying What Works for Struggling Readers: Educator's Guide*. Best Evidence Encyclopedia. Center for Data-Driven Reform in Education, Johns Hopkins University. Retrieved from www.bestevidence.org.

Smith, T., Cobb, P., Farran, D., Cordray, D. & Munter, C. (2013). Evaluating Math Recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement." *American Educational Research Journal, 50*(2), 397-428.

Stallings, J., (1980). *Allocated academic learning time revisited, or beyond time on task*. *Educational Researcher, 8*(11).

Thomas, G. & Dowker, A. (2000). *Mathematics anxiety and related factors in young children*. Paper presented at British Psychological Society Developmental Section Conference, Bristol, September 15th, 2000.

Visher, M. G., Weiss, M. J., Weissman, E., Rudd, T., and Wathington, H. D. (2012). *The effects of learning communities for students in developmental education: A synthesis of findings from six community colleges.* New York: National Center for Postsecondary Research, Teachers College, Columbia University.

Vadasy, P. F., & Sanders, E. A. (2008). Individual tutoring for struggling readers: Moving research to scale with interventions implemented by paraeducators. In G. A. Reid, A. Fawcett, F. Manis, & L. Seigel (Eds.), *The SAGE Handbook of Dyslexia* (pp. 337-355). Thousand Oaks, CA: Sage Publications, Inc.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology, 98*(3), 508-528.

Wasik, B. (2008). When fewer is more: Small groups in early childhood classrooms. E*arly Childhood Education Journal, 35*, 515-521.

Webb, N. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*(5), 366-389.

Weiss, M.J., Lockwood, J.R., & McCaffrey, D. F. (2014). Estimating the standard error of the impact estimator in individually randomized trials with clustering. *MDRC Working Papers on Research Methodology*. MDRC.

What Works Clearinghouse. (2017). *What Works Clearinghouse standards handbook* (v. 4.0). Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

White, W. B. (2015). *The Relationship Between an Affective Instructional Design, Children's Attitudes Toward Mathematics, and Math Learning for Kindergarten-Age Children* (Doctoral dissertation). Retrieved from http://dc.etsu.edu/etd/2554

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology, 5,* Article 213.

Yackel, E., Cobb, P., & Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education, 22*(5), 390-408.

Table 1: Descriptive statistics for High 5s facilitators (N=24)

Characteristic		Min	Max
Race/ethnicity (%)			
Hispanic	29.2		
Non-Hispanic white	37.5		
Non-Hispanic black	29.2		
Other/multiracial[a]	4.2		
Bachelor's degree (%)	83.3		
Degree in education (%)	26.1		
Female (%)	75.0		
Fluent in Spanish (%)	25.0		
Age (mean)	26.3	22	39
Years of formal teaching experience (mean)	1.6	0	5.8
Years of informal experience working with kids (mean)	4.0	0	14

Notes: Demographics were collected for the 24 full-time facilitators who remained at the end of the program.

[a]Other/multiracial includes Asian, Native Hawaiian/Pacific Islander, and American Indian/Alaska Native; facilitators who identified as the option "some other race"; and facilitators who selected more than one race on the survey.

Figure 1: Hypothesized pathways of influence connecting the High 5s math clubs program to student achievement.

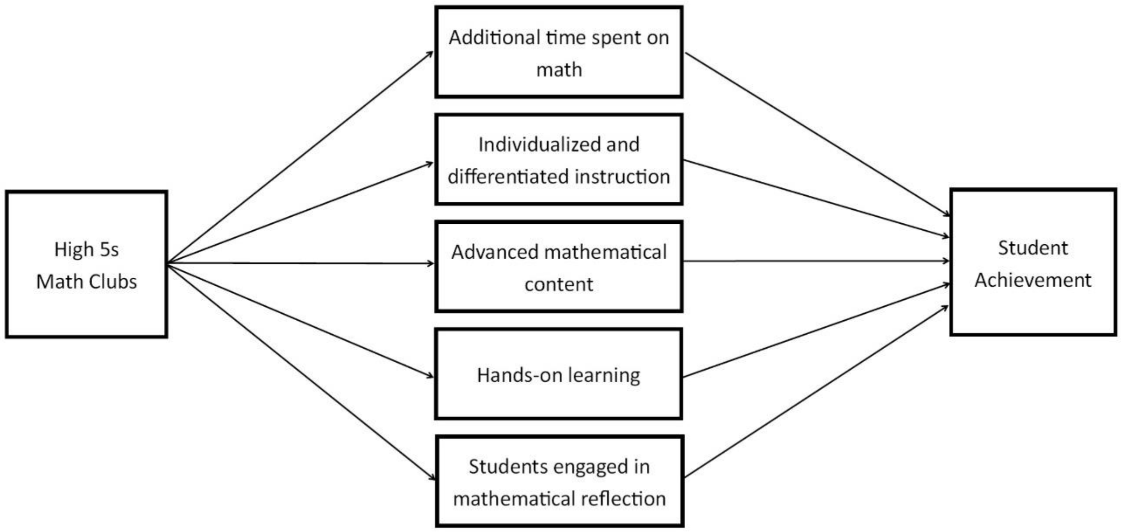

Table 2: Baseline (spring of pre-kindergarten) descriptive statistics for High 5s study sample

Student characteristics	Treatment (n=303)	Control (n=310)
Parent demographics		
Race and ethnicity (%)		
Hispanic	0.51	0.54
Non-Hispanic white	0.05	0.06
Non-Hispanic black	0.38	0.33
Other/multiracial[a]	0.06	0.05
Highest level of education		
At least high school/GED (%)	0.77	0.74
Child demographics		
Average age in fall 2015 (years)	5.17	5.17
Female (%)	0.56	0.49
English-speaking (%)	0.90	0.87
Child skills at the end of pre-K (mean)		
Math		
ECLS-B math score (0-44)	27.33	26.99
Woodcock-Johnson Applied Problems Standard Score	102.70	101.89
Language		
ROWPVT Standard Score	97.04	96.50
Executive function		
Pencil Tap: proportion correct (0-1)	0.76	0.74
Arrows Mixed: proportion correct (0-1)	0.83	0.79
Corsi Blocks forward: number correct	2.96	2.99
PSRA Attention and Inhibition Score (0-3)	2.68	2.67
Joint test of difference between groups	(F-value = 0.71)	

Notes: Rounding may cause slight discrepancies in sums and differences.
[a]Other includes Asian, Native Hawaiian/Pacific Islander, and American Indian/Alaska Native.

Figure 2: Pattern of attrition for child assessment data

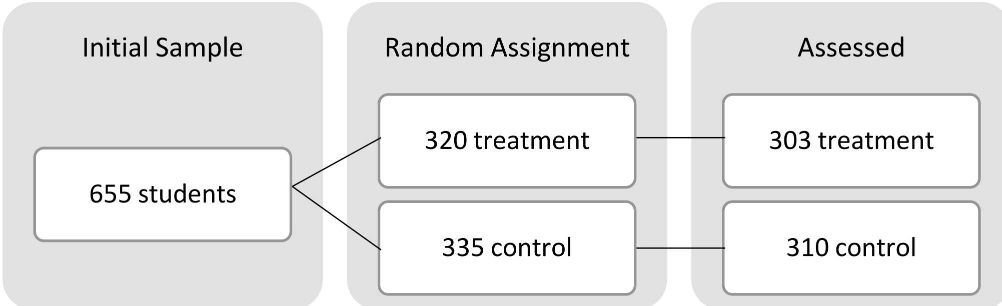

Table 3: Measures of structural fidelity based on facilitator logs and club observations

Measures of Structural Fidelity (%)	Facilitator Logs	Club Observations Fall	Club Observations Spring
Overall child attendance rate[a]	87.0	-	-
Scheduled session held	92.5	-	-
Completed at least one start-up	98.6	96.0	100.0
Completed both start-ups	89.9	92.0	92.3
Completed main activity	99.6	100.0	100.0
Correct activities conducted	96.2	96.0	92.3
Facilitator met or exceeded expectations [b]			
Conducted activities as written	-	96.0	92.3
Set up materials correctly/was familiar with the activity	-	88.0	100.0
Displayed an understanding of mathematical concepts, using correct vocabulary, and making no significant mathematical mistakes	-	100.0	96.2
Sample size			
Sites	24	24	23
Club sessions [c]	5,971	25	26

NOTES:
[a]Attendance rates are based the total number of children expected to attend each club session. Students who left the school or dropped out of clubs were not included in attendance totals after they dropped out or left the school.
[b] Facilitators were rated on a 1 to 5 scale. Facilitators who met or exceeded expectations received a rating of 3 or above.
[c] For facilitator logs, the sample size for club sessions is the total number of expected sessions. The denominator for the variables related to implementation fidelity from the facilitator logs includes only sessions with scheduled activities (5,971), not Game Days, when students were given the opportunity to choose from four activities set out by the facilitator.

Table 4. Amount of time spent on math in High 5s Clubs, based on facilitator logs and spring club observations

	Facilitator Logs		Spring Club Observations	
Use of Time	Mean	Standard Deviation	Mean	Standard Deviation
Minutes spent on start-up activities	8.4	2.26	8.5	2.42
Minutes spent on main activities	16.5	3.03	15.6	2.95
Minutes spent on math [a]	24.7	4.17	24.1	4.11
Sample size				
Sites	24		23	
Clubs	79		26	
Club sessions [b]	5,971		26	

NOTES: [a]Minutes spent on math includes time spent on start-up and main activities.
[b] The number of club sessions includes only those with scheduled activities and does not include Game Days, when students were given the opportunity to choose from four activities set out by the facilitator. During observation, only one session of each club was observed, so the number of sessions observed is the same as the number of clubs.

Table 5. Measures of process fidelity based on club observations

Quality Measure	Fall Average	Spring Average
Facilitator has good rapport with students	4.2	4.3
Facilitator avoids negative classroom management strategies	4.3	4.4
Facilitator makes math learning fun	3.6	3.7
Facilitator uses positive classroom management strategies	4.0	3.8
Facilitator encourages effort and persistence over right answer	3.4	3.2
Facilitator draws attention to the math	3.4	3.8
Facilitator underscores the mathematical objective	3.4	3.6
Facilitator encourages mathematical reflection	3.2	3.2
Facilitator asks open-ended questions	3.2	2.7
Facilitator provides opportunity for all to participate	3.9	3.8
Sample size		
Clubs	25	26

Notes: Facilitators were rated on a 1 to 5 scale. Facilitators who met or exceeded expectations received a rating of 3 or above. Facilitators who exceeded expectations received a rating of 4 or above.

Table 6: Comparison of mathematics instruction in K classrooms and High 5s Clubs

Outcome Measure	Kindergarten Average [a]	High 5s Average [b]
Time on math (min/week)	280	75
Percentage of math time in small groups	5	100
Percentage of math time using materials other than worksheets	35	100
Teacher/facilitator asks open-ended questions (1-5, where 3=satisfactory)	2.0	2.7
Teacher/facilitator encourages mathematical reflection (1-5, where 3=satisfactory)	2.7	3.2
Teacher/facilitator scaffolds children to help them extend their math skills (1-5, where 3=satisfactory)	2.5	2.9
Teacher/facilitator changed math materials/content based on individual child skill level (% of observations)	35.9	57.7
Sample size		
Sites	24	23
Classrooms	39	-
Clubs	-	26

Notes: [a] This calculation takes into account all formal math activities that occurred in the math block observations and only those that occurred during the math block portion of the full-day observations. The time-weighted average is calculated by multiplying the quality rating by the duration in minutes for each formal math activity in a classroom and then averaging across classrooms.
[b] Quality ratings were measured for the whole High 5s club which included one main activity and two start-ups. Rating are for spring only.

Table 7: Comparison of mathematics content in K classrooms and High 5s clubs

Math Activities (%)	Kindergarten Fall Observations	Kindergarten Spring Observations	High 5s Activities School Year
Numbers and operations	94.3	91.3	79.0
Numeral recognition and writing	8.3	4.4	0.0
Counting forward by 1s	25.5	21.2	0.0
Recognizing quantity without counting	1.3	0.0	3.1
Complex counting	1.3	6.6	24.6
Comparing and ordering	24.2	6.6	11.3
Composing numbers	28.0	13.9	22.6
Adding and subtracting	5.7	38.7	17.4
Other math content areas	5.7	8.8	21.0
Patterning	1.9	0.0	4.6
Shapes	2.6	7.3	11.8
Measurement	0.0	0.0	4.6
Other	1.3	1.5	-
Sample size			
Sites	21	24	-
Classrooms	36	39	-
Math activities [a]	157	137	195

Notes: [a] In kindergarten classrooms, a math activity is defined as a formal math activity observed using the COEMET, which must meet the following criteria: (1) persists for at least 1 minute, (2) develops mathematics knowledge, and (3) has a discernible topic, goal, and task. All formal math activities observed in the math block observations are included. In the full-day observations, only formal math activities taking place during regularly scheduled math instruction time are included. Classroom observation data were also verified against the full-scale GoMath! curriculum in use in most classrooms. In High 5s, a math activity is defined as either a start-up or main activity as outlined in the High 5s Full-Scale Curriculum.

Table 8: Impacts of the High 5s math clubs on student achievement

Outcome Variable	Program Group Mean (n=303)	Control Group Mean (n=310)	Difference (Impact)	P-Value	Effect Size
Math					
REMA-K[b]	39.47	37.90	1.57	0.01**	0.19
Woodcock-Johnson Applied Problems	104.17	102.98	1.19	0.18	0.09
Math attitudes					
Children's attitudes toward math (1-5)	3.51	3.45	0.07	0.58	0.04
Language					
Receptive vocabulary	97.89	97.03	0.85	0.48	0.06
Executive function					
Inhibitory control (0-1)	0.68	0.68	-0.01	0.69	-0.03
Working memory	2.29	2.22	0.07	0.51	0.05

NOTES: Statistical significance levels are indicated as follows: *** = 1 percent; ** = 5 percent; * = 10 percent. Statistically significant differences in impact estimates across different subgroups are indicated as follows: ††† = 1 percent; †† = 5 percent; † = 10 percent.

The program group received Making Pre-K Count (MPC) in pre-K and High 5s in kindergarten. The control group received only MPC in pre-K and participated in kindergarten as usual.

Impacts were estimated by comparing kindergarten outcomes for the group assigned to High 5s in kindergarten with corresponding outcomes for the group assigned to kindergarten as usual, with an adjustment for selected background characteristics and dummy variables for pre-K sites.

Outcomes were measured by T scores from the Research-Based Early Math Assessment–Kindergarten (REMA-K; Clements, Sarama, and Liu, 2008); the Woodcock-Johnson Applied Problems subscale of the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, and Mather, 2001); an MDRC-created assessment measuring children's attitudes toward math and school; the Receptive One-Word Picture Vocabulary Test (ROWPVT-4; Martin and Brownell, 2011); the Hearts and Flowers (Wright and Diamond, 2014) computerized task; and the Corsi Blocks task (Corsi, 1972; Lezak, 1983).

Rounding may cause slight discrepancies in sums and differences.

Missingness was less than 5% for most covariates but was as high as 50% for some pretest variables because only a random sample of students in the sample had been selected for testing in the spring of their pre-K year. A missingness dummy was included for all missing variables. Mean imputation was conducted with 50 imputations.

Table 9: Impacts of the High 5s Supplement in the Spring of the Kindergarten Year, by Baseline Inhibitory Control

	Low Inhibitory Control[a]				High Inhibitory Control[b]					
Outcome Score	Control Group Mean	Difference (Impact)	P-Value	Effect Size[c]	Control Group Mean	Difference (Impact)	P-Value	Effect Size[c]	Difference between subgroups	P-value
Math										
REMA-K[d]	35.06	1.26	0.37	0.15	41.84	0.63	0.34	0.11	0.63	0.68
Woodcock-Johnson Applied Problems[e]	99.63	0.67	0.72	0.05	106.99	1.98*	0.09*	0.17	-1.31	0.55
Math attitudes										
Children's attitudes toward math (1-5)	3.51	0.03	0.92	0.02	3.55	0.07	0.75	0.04	-0.04	0.92
Language										
Receptive vocabulary[e]	97.49	-3.44	0.26	-0.24	102.51	0.24	0.88	0.02	-3.68	0.29
Executive function										
Inhibitory control (0-1)	0.63	0.01	0.82	0.05	0.71	0.00	0.97	0.01	0.88	0.01
Working memory[f]	2.16	-0.17	0.49	-0.14	2.62	0.00	0.98	-0.16	0.58	-0.16
Sample size										
Sites	22				24					
Children	57				99					

NOTES: Statistical significance levels are indicated as follows: *** = 1 percent; ** = 5 percent; * = 10 percent. Statistically significant differences in impact estimates across different subgroups are indicated as follows: ††† = 1 percent; †† = 5 percent; † = 10 percent.

Impacts were estimated by comparing kindergarten outcomes for the group assigned to High 5s in kindergarten with corresponding outcomes for the group assigned to kindergarten as usual, with an adjustment for selected background characteristics and dummy variables for pre-K sites.

Outcomes were measured by T scores from the Research-Based Early Math Assessment–Kindergarten (REMA-K; Clements, Sarama, and Liu, 2008); the Woodcock-Johnson Applied Problems subscale of the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, and Mather, 2001); an MDRC-created assessment measuring children's attitudes toward math and school; the Receptive One-Word Picture Vocabulary Test (ROWPVT-4; Martin and Brownell, 2011); the Hearts and Flowers (Wright and Diamond, 2014) computerized task; and the Corsi Blocks task (Corsi, 1972; Lezak, 1983).

Rounding may cause slight discrepancies in sums and differences.

Baseline inhibitory was only available for a random subset of study participants, thus the total sample size for this analysis is smaller than for the main impacts. Missingness was less than 5% for most covariates but was as high as 50% for some pretest variables because only a random sample of students in the sample had been selected for testing in the spring of their pre-K year. A missingness dummy was included for all missing variables. Mean imputation was conducted with 50 imputations.

Table 10: Impacts of the High 5s Supplement in the Spring of the Kindergarten Year, by Math Instructional Time

	Low Kindergarten Math Instructional Time[a]				High Kindergarten Math Instructional Time[b]					
Outcome Score	Control Group Mean	Difference (Impact)	P-Value	Effect Size[c]	Control Group Mean	Difference (Impact)	P-Value	Effect Size[c]	Difference Between Subgroups	P-Value
Math										
REMA-K[d]	39.88	0.82	0.36	0.10	36.15	2.33	0.01***	0.29	-1.51	0.23
Woodcock-Johnson Applied Problems[e]	103.68	1.92	0.17	0.14	102.33	0.70	0.56	0.06	1.22	0.51
Math attitudes										
Children's attitudes toward math (1-5)	3.28	0.44	0.01**	0.29	3.59	-0.25	0.13	-0.18	0.69	0.00†††
Language										
Receptive vocabulary[e]	98.71	2.09	0.24	0.13	95.78	-0.23	0.88	-0.02	2.32	0.33
Executive function										
Inhibitory control (0-1)	0.73	-0.03	0.26	-0.13	0.65	0.02	0.50	0.08	-0.05	0.20
Working memory[f]	2.41	0.08	0.63	0.05	2.04	0.09	0.52	0.07	-0.02	0.94
Sample size										
Sites	11				13					
Children	142				168					

NOTES: Statistical significance levels are indicated as follows: *** = 1 percent; ** = 5 percent; * = 10 percent. Statistically significant differences in impact estimates across different subgroups are indicated as follows: ††† = 1 percent; †† = 5 percent; † = 10 percent.

The program group received Making Pre-K Count (MPC) in pre-K and High 5s in kindergarten. The control group received only MPC in pre-K and participated in kindergarten as usual.

Impacts were estimated by comparing kindergarten outcomes for the group assigned to High 5s in kindergarten with corresponding outcomes for the group assigned to kindergarten as usual, with an adjustment for selected background characteristics and dummy variables for pre-K sites.

Outcomes were measured by T scores from the Research-Based Early Math Assessment–Kindergarten (REMA-K; Clements, Sarama, and Liu, 2008); the Woodcock-Johnson Applied Problems subscale of the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, and Mather, 2001); an MDRC-created assessment measuring children's attitudes toward math and school; the Receptive One-Word Picture Vocabulary Test (ROWPVT-4; Martin and Brownell, 2011); the Hearts and Flowers (Wright and Diamond, 2014) computerized task; and the Corsi Blocks task (Corsi, 1972; Lezak, 1983).

Rounding may cause slight discrepancies in sums and differences.

Missingness was less than 5% for most covariates but was as high as 50% for some pretest variables because only a random sample of students in the sample had been selected for testing in the spring of their pre-K year. A missingness dummy was included for all missing variables. Mean imputation was conducted with 50 imputations.

[a]Children in schools where the average duration of math instructional time recorded during observations of kindergarten classrooms was less than 53 minutes constitute the low kindergarten math instructional time group.

[b]Children in schools where the average duration of math instructional time recorded during observations of kindergarten classrooms was greater than or equal to 53 minutes constitute the high kindergarten math instructional time group.

[c]Effect size is calculated by dividing the impact of the program (the difference between the means for the program group and the control group) by the standard deviation for the control group.

[d]The REMA-K has a mean score of 50 and a standard deviation of 10. The norm is based on a group of children between pre-K and third grade.

[e]This is a standardized measure with a mean score of 100 and a standard deviation of 15.

[f]The score reports the highest number of blocks the child was able to tap in correct order in two attempts.